

ETHICS IN MACHINE LEARNING

Sometime in the Future

Instructor:	Anqi Liu	Time:	Some perfect time
Email:	XYZ@email.org	Place:	Some perfect location
Office Hour:	(Twice a week)		

Objectives: This course is primarily designed for graduate students who already has a basic idea of modern machine learning techniques. By the end of this course, you will be able to:

- recognize the importance of ethical considerations in design machine learning algorithms and data science solutions to real-world problems;
- describe one aspect of your interest in a range of topics surrounding ethical consideration of algorithm design;
- investigate and critique the current methods in the chosen area;
- apply recent advanced techniques in the area of fairness, privacy, and safety to real world problems;
- design or implement your own solution to one problem with your consideration in mind.

Prerequisites: An undergraduate-level understanding of probability, statistics, algorithms, and linear algebra is assumed. Experience of graduate level machine learning research preferred.

Tentative Course Outline: (aligned with assessments)

- Week1: Class policy. Motivation. Introduction of prepared topics.
- Week2: Introduction of prepared topics cont', brainstorming additional topics. Full paper list released.
- Week3: Statement of Interest due. Fairness in ML (I).
- Week4: Statement of Interest feedback. Fairness in ML (II).
- Week5: Paper choice and group assignment due. Differential Privacy (I).
- Week6: Paper choice feedback. Differential Privacy (II).
- Week7: Discussion class (I). Peer feedback on the on-going paper critique.
- Week8: Paper Critique due. AI safety (I).
- Week9: Discussion class (II). Work out a project proposal together: what do you want to achieve and how can you design your own solution?
- Week10: Project proposal due. AI safety (II).
- Week11: Proposal feedback. Additional topics that are of interest (I).
- Week12: Additional topics that are of interest (II).
- Week13: Presentation.

Grading: (aligned with learning outcome, focus on formative assessment)

- Statement of interest: state what and why you think the proposed aspect is important to real-world deployment, list at least one application that is associated with this aspect. (10%)
- Paper critique: choose one paper/article under the proposed topic and discuss the pros and cons of one current method developed under ethical principles, as well as difficulties and challenges that still exist. (20%)

- Class attendance (10%) and Discussion class participation (15%).
- Project proposal (each group): propose your own solution, plan out timeline and task assignment. Students will be encouraged to apply learned techniques in class, but are also welcomed to explore novel methods. (20%)
- Project presentation (each group): present your solution in class. (25%)

Course Pages:

1. https://webpage_placeholder.com
2. Resource list (click me!): <https://docs.google.com/document/d/1dPV8J4MVnzUiLjzoNJDBTkfgFbzwkZfeIj4PDE/edit?usp=sharing>

Suggested Books:

1. The Ethical Algorithm, <https://www.amazon.com/Ethical-Algorithm-Science-Socially-Design/dp/0190948205>
2. Fairness and Machine Learning, <https://fairmlbook.org/>
3. The Algorithmic Foundations of Differential Privacy, <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>
4. Human Compatible: AI and the Problem of Control, <https://www.amazon.com/Human-Compatible-Artificial-Intelligence/dp/B07N5J5FTS>

Important Dates:

- Red on the schedule: Due dates (Important!)
- Blue on the schedule: Discussion class (Don't miss;)).

Course Policy: (encouraging, positive, and respectful)

- Discussion and collaboration is encouraged in the projects.
- Group assignment guidelines: After the submission of the statement of interest, the students are encouraged to find other students who are interested in similar topics (for example, they are all interested in fairness). In the feedback of paper choice and group assignment, the instructor will try to coordinate students with similar interests and different paper choices in the same group.
- Discussion class I: This discussion class is designed to facilitate the communication of different chosen topics and papers by the students. The students are expected to give peer feedback to each other about their chosen paper and on-going critique. The students will be divided into groups (other than the project groups) and will be asked to report to the class about one most unexpected/difficult ethical consideration from the discussion.
- Discussion class II: This discussion class is designed to facilitate the consensus of each group in the project proposal. The students will stay in the project group. The students will choose one application from the paper critiques they are working on and propose their solution. At the end of the class, the students will be asked to report to the class their choice and reasons.

Class Policy:

- Regular attendance is essential and expected.